NCEI v2.0 timeSeries Example netCDF File Testing Report

Mathew Biddle
NOAA's National Centers for Environmental Information (NCEI)
Cooperative Institute for Climate and Satellites - Maryland
Silver Spring, MD

This document provides a brief summary of using the NCEI netCDF templates v2.0 in various online compliance checkers, software packages, and web services. The intention of this report is to provide further detail concerning the various messages a netCDF file would receive if it was in complete compliance with the NCEI netCDF templates v2.0 for the timeSeries featureType. In addition, the example data files have been imported into various software packages (Matlab, Python, and ODV) to document any inconsistencies when ingesting and plotting the data points. The commands used to import and plot the data in the software packages have also been included.

The data files used in this report can be found at:
http://data.nodc.noaa.gov/ncei/example/data/netcdf/v2.0/
 and in the THREDDS catalog:
http://data.nodc.noaa.gov/thredds/catalog/example/v2.0/catalog.html

Version control

Version	Date	Author	Comments
0.9	2016-09-30	M. Biddle	Initial draft of documentation.
1.0	2016-10-11	M. Biddle	Updated documentation with feedback from NCEI netCDF working group.

Below is an outline of how this document is constructed:

Template Feature Type: (the feature type used in testing, hyperlink to file)

Checker: (compliance checker where the issue arose, hyperlink to checker)

Specific check: (includes score and date the check was ran on)

- **Issue #1**:
  - Possible resolution
- **Issue #2:**
  - Possible resolution

Specific check:

- **Issue #1**:
  - Possible resolution
- **Issue #2:**
  - Possible resolution

Checker:

Specific check:

- **Issue #1**:
  - Possible resolution
- **Issue #2:**
  - Possible resolution

Specific check:

- **Issue #1**:
  - Possible resolution
- **Issue #2:**
  - Possible resolution

Access Site: (the access site used for the protocols below, e.g. THREDDS)

Protocol: (the access protocol tested, e.g. OPeNDAP Dataset Access Form)

- Result of test

Protocol:

- Result of test

Software Package: (name of software package, e.g. Matlab)

- **Opening/Importing the file.**
  - Result of the test, including code.
- **Plotting the data**
  - Result of test, including code and plots.

timeSeries:

IOOS Compliance Checker:

CF 1.6 (157/165) ran on 09/19/2016:

- **Cell_methods_name: The name field does not match a dimension, area or coordinate.**
  - The contents of the cell_methods attribute should be "lon: point lat: point" instead of "longitude: point latitude: point".
  - CF 1.6 section 7.3.4 states that it is fine for a cell method to reference a standard name (such as longitude) even though there is no coordinate designated as longitude, but it does not allow for naming a dimension. It also allows for using the name 'area', but it forbids using a standard name or 'area' if that name corresponds to a scalar coordinate variable name or a dimension name that doesn't have a corresponding 1D coordinate variable of the same name.
  - The ability to use a standard name is important for cases where each cell represents an integral over a coordinate that is not present in the file, such as Hovmöller diagram grids (time vs latitude or longitude).
- **The variable instrument1 does not have associated coordinates**
  - Expected, as the instrument variable doesn't need to have coordinates.
- **The variable platform1 does not have associated coordinates**
  - Expected, as the platform variable doesn't need to have coordinates.
- **Variable sal has a non-space-time dimension after space-time-dimensions**
  - §2.4 Dimension order - "All other dimensions should, whenever possible, be placed to the left of the spatiotemporal dimensions."
  - File is sal(timeSeries,time).
  - This is an error in the checker, the developers are aware of the problem (https://github.com/ioos/compliance-checker/issues/284) and are working to resolve it (https://github.com/ioos/compliance-checker/milestone/8).
- **Variable temp has a non-space-time dimension after space-time-dimensions**
  - §2.4 Dimension order - "All other dimensions should, whenever possible, be placed to the left of the spatiotemporal dimensions."
  - File is temp(timeSeries,time).
  - This is an error in the checker, the developers are aware of the problem (https://github.com/ioos/compliance-checker/issues/284) and are working to resolve it (https://github.com/ioos/compliance-checker/milestone/8).

ACDD 1.3 (97/107) ran on 09/19/2016:

- **Neither 'acknowledgment' nor 'acknowledgement' attributes present**
  - Ignore. File contains "acknowledgement"
  - This is an error in the checker and has been changed, it has yet to be published in the web checker (https://github.com/ioos/compliance-checker/pull/262).
- **Var instrument1 missing attr standard_name and units**

- ○ Expected error, instrument variable doesn't need to have standard_name and units.
- ○ There is no requirement for any variable to have a standard name. If the quantity being measured doesn't have one it doesn't have one. This should be treated as informational.
- **Var platform1 missing attr standard_name and units**
  - ○ Expected error, platform variable doesn't need to have standard_name and units.
  - ○ There is no requirement for any variable to have a standard name. If the quantity being measured doesn't have one it doesn't have one. This should be treated as informational.
- **Var timSeries missing attr coverage_content_type, standard_name, and units**
  - ○ Expected error, profile variable doesn't need to have coverage_content_type, standard_name and units.
  - ○ Since it is a container variable, it does not need to have those attributes.
- **Attr geospatial_lat_resolution not present**
  - ○ Since it's only one point in space, the resolution attribute is not necessary.
- **Attr geospatial_lon_resolution not present**
  - ○ Since it's only one point in space, the resolution attribute is not necessary.
- **Attr geospatial_vertical_resolution not present**
  - ○ Since it's only one point in space, the resolution attribute is not necessary.

NCEI Timeseries Orthogonal 1.1 (115/124) ran on 09/19/2016

- **nodc_template_version attribute must be NODC_NetCDF_TimeSeries_Orthogonal_Template_v1.1**
  - ○ This is an expected error, the file contains NCEI_NetCDF_TimeSeries_Orthogonal_Template_v2.0 in the nodc_template_version attribute which is in compliance with current practices.
- **Conventions attribute is missing or is not equal to CF-1.6: CF-1.6, ACDD-1.3**
  - ○ This is an expected error, current conventions allow for multiple conventions attributes.
  - ○ This is a problem in the checker and is currently being revisited, see https://github.com/ioos/compliance-checker/milestone/8 for more information.
- **Metadata_Conventions attribute is required to be 'Unidata Dataset Discovery v1.0'**
  - ○ This is an expected error, ACDD 1.3 changed the recommendation for this attribute.
- **geospatial_lat_resolution should exist and not be empty.**
  - ○ This is an expected error, resolution attributes do not apply to single point timeSeries data sets.
- **geospatial_lon_resolution should exist and not be empty.**
  - ○ This is an expected error, resolution attributes do not apply to single point timeSeries data sets.
- **geospatial_vertical_resolution should exist and not be empty.**
  - ○ This is an expected error.
  - ○ Resolution attributes do not apply to single point timeSeries data sets.
- **platform should exist and point to a variable.**
  - ○ This is an expected error.
  - ○ Platform now contains ACDD representation of the platform.

- **sea_name attribute should exist and should be from the NODC sea names list: Cordell Bank National Marine Sanctuary is not a valid sea name**
  - Checker is looking at https://github.com/lukecampbell/cc-plugin-ncei/blob/base/cc_plugin_ncei/data/seanames.csv for seanames. it only includes those with IHB codes. Thus, it will never include "Coastal Waters" or "Marine Sanctuaries"
  - The problem has been reported at https://github.com/ioos/cc-plugin-ncei/issues/8 and is being investigated. This check can be ignored.

Jet Propulsion Laboratory Metadata Compliance Checker:

ACDD 1.3 Checker (107/147):

- **"There was a problem downloading your file. The server may be too busy or the file may not exist. URL: http://data.nodc.noaa.gov/thredds/dodsC/testdata/mbiddle/GOLD_STANDARD_NETCDF/2.0/NCEI_timeSeries_template_v2.0_2016-06-15_133831.473835.nc"**
  - Issue with the checker using OPeNDAP links.
  - Downloading the file and testing against that file works.
  - Bug with the checker.
- **check for a comma separated value failed because "keywords" might not be comma separated**
  - Ignore as keywords is comma separated, "keywords: Oceans > Ocean Temperature > Water Temperature, Oceans > Salinity/Density > Salinity"
  - The list in the keywords attribute uses the GCMD convention for describing keywords, to find more information about GCMD keywords, see http://gcmd.nasa.gov/learn/keywords.html.
- **check for a comma separated value failed because "Conventions" might not be comma separated**
  - Ignore as conventions is comma separated, "Conventions: CF-1.6, ACDD-1.3"
- **check for existence failed because "geospatial_*_resolution" does not exist**
  - Ignore as point is only one lat/lon/z instance so resolution is moot.
- **check for existence failed because "crs:long_name" does not exist**
  - Ignore because crs is not required to have long_name attribute.
- **standard_name 12 out of 20 passed**
  - Not all variables require a standard_name attribute, in this case instrument1, platform1, and crs are not required to have standard_name attributes. So the standard_name checks are not required for those variables.
- **units 12 out of 20 passed**
  - Not all variables require a units attribute, in this case instrument1, platform1, and crs are not required to have units attributes. So the units checks are not required for those variables.
- **coverage_content_type 4 out of 20 passed**
  - Not all variables require a coverage_content_type attribute, in this case instrument1, platform1, crs, z, time, lat, and lon are not required to have content_coverage_type

attributes. So the content_coverage_type checks are not required for those variables. Content_coverage_type only applies to geophysical variables.

CF 1.6 Checker (159/178):

- **lat is not allowed to have an axis attr as it is not a coordinate var**
  - Our templates define this variable as an auxiliary coordinate variable, since the variable does not have the same name as the dimension. See http://www.unidata.ucar.edu/software/netcdf/workshops/2011/datamodels/NcCVars.html for more information on coordinate variables.
  - According to CF conventions, auxiliary coordinate variables are not allowed to have axis attributes.
  - This is an expected result and is an error in our templates. The variables lat, lon, and z should not have axis attributes.
- **lon is not allowed to have an axis attr as it is not a coordinate var**
  - Our templates define this variable as an auxiliary coordinate variable, since the variable does not have the same name as the dimension. See http://www.unidata.ucar.edu/software/netcdf/workshops/2011/datamodels/NcCVars.html for more information on coordinate variables.
  - According to CF conventions, auxiliary coordinate variables are not allowed to have axis attributes.
  - This is an expected result and is an error in our templates. The variables lat, lon, and z should not have axis attributes.
- **z is not allowed to have an axis attr as it is not a coordinate var**
  - Our templates define this variable as an auxiliary coordinate variable, since the variable does not have the same name as the dimension. See http://www.unidata.ucar.edu/software/netcdf/workshops/2011/datamodels/NcCVars.html for more information on coordinate variables.
  - According to CF conventions, auxiliary coordinate variables are not allowed to have axis attributes.
  - This is an expected result and is an error in our templates. The variables lat, lon, and z should not have axis attributes.
- **Conventions field is not "CF-1.6"**
  - Ignore this error as "Conventions: CF-1.6, ACDD-1.3"
- **Units attribute required for timeSeries variable**
  - Ignore, timeSeries variable does not need to have units.
- **Sal units are 0.001, standard_name units should be 1e-3**
  - Ignore, 0.001 = 1e-3
- **Temp units are degree_Celsius, standard_name units should be K**
  - Ignore, degree_Celsius is a UDUNIT and is sufficient in this context.
- **The timeSeries dimension for the variable lat does not have an associated coordinate variable, but is a Lat/Lon/Time/Height dimension.**
  - Technically, no coordinate variables exist in the file. (according to http://www.unidata.ucar.edu/software/netcdf/workshops/2011/datamodels/NcCVars.html)
  - The checker wants to find a variable that has the same name as a dimension, however, for a point data set this is unreasonable. We identified one dimension which gets used as the primary dimension.

- **The timeSeries dimension for the variable lon does not have an associated coordinate variable, but is a Lat/Lon/Time/Height dimension.**
  - Technically, no coordinate variables exist in the file. (according to http://www.unidata.ucar.edu/software/netcdf/workshops/2011/datamodels/NcCVars.html)
  - The checker wants to find a variable that has the same name as a dimension, however, for a point data set this is unreasonable. We identified one dimension which gets used as the primary dimension.
- **Sal cell_methods_name all 3 failed "The name field does not appear in the allowable types.","The name field does not match the dimension.","The name field does not match the reserved words "interval", "area", or "comment"."**
  - "longitude: point latitude: point" should be "lon: point lat: point"
- **The timeSeries dimension for the variable sal does not have an associated coordinate variable, but is a Lat/Lon/Time/Height dimension.**
  - Technically, no coordinate variables exist in the file. (according to http://www.unidata.ucar.edu/software/netcdf/workshops/2011/datamodels/NcCVars.html)
  - The checker wants to find a variable that has the same name as a dimension, however, for a point data set this is unreasonable. We identified one dimension which gets used as the primary dimension.
- **Temp cell_methods_name all 3 failed "The name field does not appear in the allowable types.","The name field does not match the dimension.","The name field does not match the reserved words "interval", "area", or "comment"."**
  - "longitude: point latitude: point" should be "lon: point lat: point"
- **The timeSeries dimension for the variable temp does not have an associated coordinate variable, but is a Lat/Lon/Time/Height dimension.**
  - Technically, no coordinate variables exist in the file. (according to http://www.unidata.ucar.edu/software/netcdf/workshops/2011/datamodels/NcCVars.html)
  - The checker wants to find a variable that has the same name as a dimension, however, for a point data set this is unreasonable. We identified one dimension which gets used as the primary dimension.
- **The timeSeries dimension for the variable timeSeries does not have an associated coordinate variable, but is a Lat/Lon/Time/Height dimension.**
  - Technically, no coordinate variables exist in the file. (according to http://www.unidata.ucar.edu/software/netcdf/workshops/2011/datamodels/NcCVars.html)
  - The checker wants to find a variable that has the same name as a dimension, however, for a point data set this is unreasonable. We identified one dimension which gets used as the primary dimension.
- **The timeSeries dimension for the variable z does not have an associated coordinate variable, but is a Lat/Lon/Time/Height dimension.**
  - Technically, no coordinate variables exist in the file. (according to http://www.unidata.ucar.edu/software/netcdf/workshops/2011/datamodels/NcCVars.html)
  - The checker wants to find a variable that has the same name as a dimension, however, for a point data set this is unreasonable. We identified one dimension which gets used as the primary dimension.

THREDDS:

OPeNDAP Dataset Access Form:

- On a Linux command line, ran:

```
ncdump -h
http://data.nodc.noaa.gov/thredds/dodsC/example/v2.0/NCEI_timeSeries_template_v2.0_2016-09-22_181608.54
8782.nc
```
- Functions as expected.

HTTPServer:

- On a Linux command line, ran:

```
wget
http://data.nodc.noaa.gov/thredds/fileServer/example/v2.0/NCEI_timeSeries_template_v2.0_2016-09-22_1816
08.548782.nc

ncdump -h NCEI_timeSeries_template_v2.0_2016-09-22_181608.548782.nc
```
- Functions as expected.

WCS:

- Functions as expected.

WMS:

- Functions as expected.

NCML:

- Functions as expected.

UDDC (43/46):

- **Missing Metadata_Conventions**
  - Attribute has been depricated in ACDD 1.3.
- **Missing geospatial_vertical_resolution**
  - Attribute not required for point data sets.
- **Missing acknowledgment**
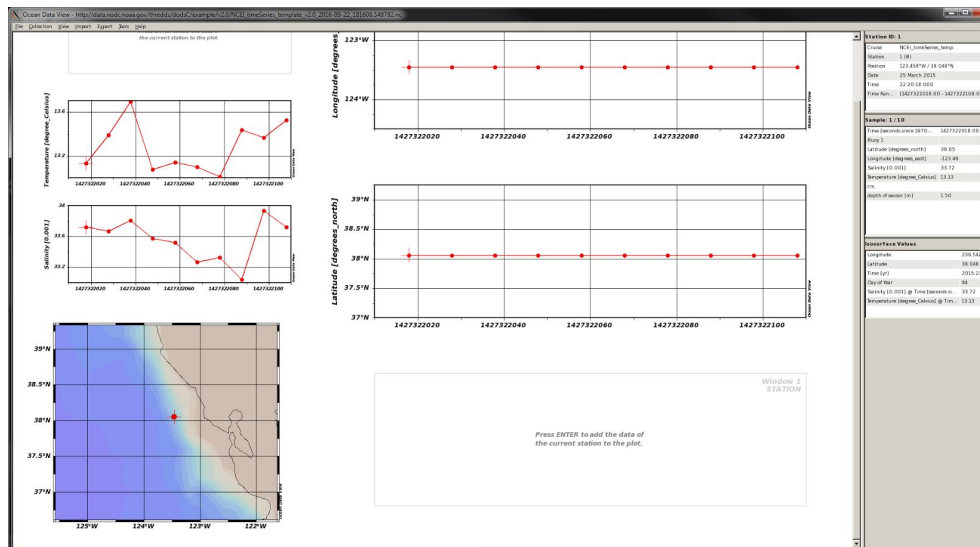 - Notice the spelling "acknowledgment" vs "acknowledgement"

ISO:

- Functions as expected.

Ocean Data View (4.5.0):

- **Open remote:**

- ○ Url:
    ```
    http://data.nodc.noaa.gov/thredds/dodsC/example/v2.0/NCEI_timeSeries_template_v2.0_2016
    -09-22_181608.548782.nc
    ```
  - ○ Select "time" and "timeSeries" dimensions, use "Time" as primary variable dimensions.
  - ○ Opens as expected
- **Plotting:**
  - ○ Plots as expected:
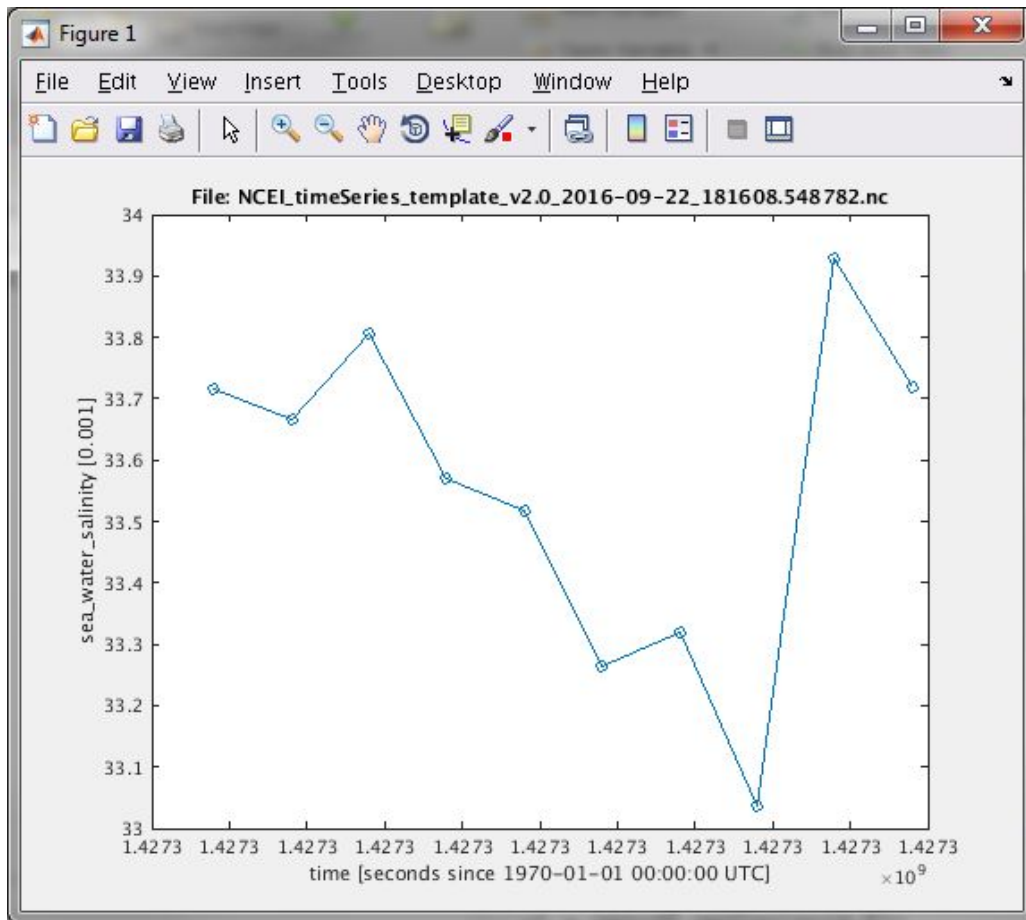

Matlab (R2015a):

- **Import the data:**
  - ○ Imports as expected:

```
url =
'http://data.nodc.noaa.gov/thredds/dodsC/example/v2.0/NCEI_timeSeries_template_v2.0_2016-09-22_
181608.548782.nc';
nc = ncinfo(url);
for i=1:length(nc.Variables) g=sprintf('%i ', nc.Variables(i).Size);
fprintf('%11s %s\n',nc.Variables(i).Name,g);
end

 timeSeries 1
       time 10
        lat 1
        lon 1
          z 1
instrument1 64
  platform1 64
        crs 1
        sal 10 1
       temp 10 1
```

- **Plotting the data:**
  - ○ Plots as expected

8

```
url =
'http://data.nodc.noaa.gov/thredds/dodsC/example/v2.0/NCEI_timeSeries_template_v2.0_2016-09-22_
181608.548782.nc';
ncid = netcdf.open(url);
x1 = netcdf.getVar(ncid,1);
y1 = netcdf.getVar(ncid,8);

plot(x1,y1,'o-')
xlabel(sprintf('%s [%s]',netcdf.getAtt(ncid,1,'standard_name'),...
    netcdf.getAtt(ncid,1,'units')),'Interpreter','none');
ylabel(sprintf('%s [%s]',netcdf.getAtt(ncid,8,'standard_name')...
    ,netcdf.getAtt(ncid,8,'units')),'Interpreter','none');
file = strsplit(url,'/');
title(sprintf('File: %s',file{end}),'Interpreter','none');
```


Python 2.6.6 (r266:84292, Aug 9 2016, 06:11:56):

- **Importing the data:**
    - Imports as expected.

```
import netCDF4
url =
'http://data.nodc.noaa.gov/thredds/dodsC/example/v2.0/NCEI_timeSeries_template_v2.0_2016-09-22_
181608.548782.nc'
f = netCDF4.Dataset(url,'r')
for v in f.variables:
```

```
    print '%11s %s %s' %(v, f.variables[v].dimensions, f.variables[v].shape)

 timeSeries (u'timeSeries',) (1,)
       time (u'time',) (10,)
        lat (u'timeSeries',) (1,)
        lon (u'timeSeries',) (1,)
          z (u'timeSeries',) (1,)
instrument1 (u'maxStrlen64',) (64,)
  platform1 (u'maxStrlen64',) (64,)
        crs () ()
        sal (u'timeSeries', u'time') (1, 10)
       temp (u'timeSeries', u'time') (1, 10)
```

- **Plotting the data:**
  - ○ Plots as expected

```
import netCDF4
import numpy as np
import matplotlib.dates as mdates
import matplotlib.pyplot as plt
import datetime

url =
'http://data.nodc.noaa.gov/thredds/dodsC/example/v2.0/NCEI_timeSeries_template_v2.0_2016-09-22_
181608.548782.nc'
f = netCDF4.Dataset(url,'r')
var1_data = f.variables['time'][:].flatten()
var2_data = f.variables['sal'][:].flatten()
map_time = map(datetime.datetime.fromtimestamp,var1_data) # map seconds since 1970 to python
date
time_num = mdates.date2num(map_time) # convert python date to number

fig = plt.figure(figsize=(15,10))
ax = fig.add_subplot(1,1,1)
ax.plot_date(x=time_num, y=var2_data,fmt='o-') # plot the timeseries
plt.xaxis.set_major_formatter(mdates.DateFormatter('%b/%y')) # format the xaxis labels
plt.xlabel('%s [%s]' %(f.variables['time'].long_name, f.variables['time'].units))
plt.ylabel('%s [%s]' %(f.variables['sal'].long_name, f.variables['sal'].units))
plt.grid(b=True, which='Major', color='k')
plt.title('File: %s' % url.split('/')[-1])
plt.show()
```
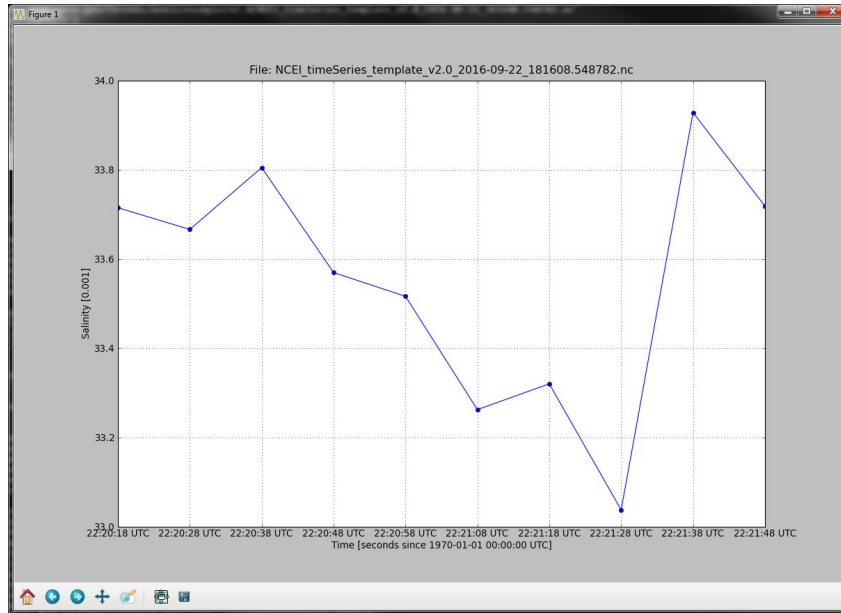
Aggregation:

- Performed a 'joinExisting' aggregation on the 'time' dimension. Although all the time epochs were aggregated, the corresponding measurements were not. THREDDS threw out an error with the index.
- Tried an alternate approach by performing a 'joinNew' aggregation on a new 'time_agg' dimension which worked.